DOCUMENT RESUME

ED 275 128 FL 015 311

AUTHOR Fodor, Janet Dean; Crain, Stephen
TITLE Simplicity and Generality of Rules in Language Acquisition.
PUB DATE 84
NOTE 13p.; Paper presented at the Annual Boston University Conference on Language Development (Boston, MA, 1984).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Child Language; Error Patterns; *Generalization; *Grammar; *Language Acquisition; *Learning Processes; *Linguistic Difficulty (Inherent); *Linguistic Theory; Morphology (Languages); Syntax

ABSTRACT

An alternative to the standard theory that language learners always formulate the simplest rule to accommodate data is proposed. This new position states that the system of formulating rules and the generalizations made from it by children and adults in the stages of language learning needs to be more specific. The present theory excludes the use of transformational and all other context-sensitive rules and is limited to the use of context-free phrase structure rules. It accepts the "slash categories" of generalized phrase structure grammar but rejects the metarules that generalize across them, because they tend to lead to the same kind of overgeneralization that a transformation leads to. A rule schema that allows rules to be added to the child's system as he learns them, by collapsing slash rules, explains piecemeal language learning and the adjustment of rules to new data presented to the child. Although problems remain in the theory, it holds promise as a basis for research into innate language acquisition and linguistic knowledge. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275128

SIMPLICITY AND GENERALITY OF RULES IN LANGUAGE ACQUISITION

Janet Dean Fodor and Spephen Crain
University of Connecticut

Boston University Conference on
Language Development, 1984

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

● This document has been reproduced as received from the person or organization originating it.

□ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Janet Fodor

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 015311

BEST COPY AVAILABLE

SIMPLICITY AND GENERALITY OF RULES IN LANGUAGE ACQUISITION

Janet Dean Fodor and Stephen Crain
University of Connecticut

Boston University Conference on
Language Development, 1984

Languages like generalizations and children like generalizations -- a perfect harmony, but of course not an accidental one. Languages are shaped by their users, particularly by their learners. It is because children generalize that languages do. We can even push the explanation back one step further. Why do children generalize? The standard answer to this rests on two assumptions: first, that language learners always formulate the simplest rule that will accommodate their data -- this is a least-effort principle which is difficult to argue with; second, that natural language grammars have the formal property that the simpler a rule is, the more general it is. In other words: children favor general rules because they are easier than restricted rules.

A very clear example of this kind of argument appears in Morris Halle's 1962 paper arguing for the psychological reality of phonological features. A rule written in features looks, on the page, more complex than a rule written in phonemes; but Halle argues that it must be the feature notation that we have inside our heads, for only this exhibits the proper POSITIVE correlation between simplicity and generality. (A rule referring to all front vowels is more general than one referring only to the front vowel /i/; in phonemes the more general rule is more complex, but in features it is simpler.)

(1)

$$a \rightarrow \quad / __ \left\{ \begin{matrix} i \\ e \\ æ \end{matrix} \right\}$$

GENERAL BUT COMPLEX

$$a \rightarrow æ \; / __ \; i$$

SPECIFIC BUT SIMPLE

(2)

$$[+grave] \rightarrow [-grave] \; / __ \begin{bmatrix} +voc \\ -cons \\ -grave \end{bmatrix}$$

GENERAL AND SIMPLE

$$[+grave] \rightarrow [-grave] \; / __ \begin{bmatrix} +voc \\ -cons \\ +diff \\ -comp \\ -flat \\ -grave \end{bmatrix}$$

SPECIFIC AND COMPLEX

Exactly comparable points have been made for syntax, leading to empirical claims about the format of syntactic rules -- e.g., that they can contain parentheses, essential variables, syntactic features, and so forth.

This is a very standard line of reasoning in linguistics, and the only problem with it is that is doesn't work. The generalizations that natural languages exhibit are PARTIAL generalizations only. In phonology there are rules that do things to consonants in the context of a front vowel or a high front vowel, but very often not in the context of just any vowel (or just any segment). But of course, if children always pick the simplest rule, and if simplicity means generality, then they should opt immediately for the 'all vowels' context rather than for the more restricted one.

(3)

PROBLEM: [+grave] --> [-grave] / ___ $\begin{bmatrix} +\text{voc} \\ -\text{cons} \end{bmatrix}$

EVEN SIMPLER THAN (2), BUT OVERGENERAL

Syntactic rules also typically express partial generalizations. In English we can topicalize a noun phrase, a prepositional phrase, an untensed verb phrase, but not a tensed verb phrase. We can topicalize a phrase out of a main clause and out of an object complement clause, but not out of a subject complement clause. We can extrapose a clause only if it has an overt complementizer; we can invert auxiliary verbs but not main verbs. And so on and so on. Here too, the straightforward story about simplicity and generality suggests that learners should favor the wrong generalizations. We would expect them always to choose the maximal generalizations, whereas in fact they hypothesize partial generalizations, and apparently just the right partial generalizations as exhibited in the adult language.

A brief word here about what we take the facts to be. Notoriously, children do overgeneralize morphological rules; they apply them to exceptional cases and produce forms like goes and mouses. They also, apparently, overgeneralize lexical rules, at least sometimes and to some extent; for example, Bowerman (1974) reports incorrect causative constructions such as This is aching my legs. But these are cases which call for some kind of brute force listing in the adult grammar. What children rarely or never do, as far as we know, is make the wrong choice between two purely syntactic generalizations. They don't construe sentences like Candy you don't eat before dinner as evidence that one can topicalize nouns, rather than noun phrases; if they did they would say things like Hat I'm gonna paint the red instead of The hat I'm gonna paint red. They also don't try topicalizing tensed verb phrases, or moving things out of

extraction islands such as subject clauses. This may be a slight exaggeration, but in what follows we will adopt the working assumption that children never overgeneralize purely syntactic rules.

There are three ways in which one could try to reconcile this fact about the errors that children don't make, with the least effort assumption about grammar development.

ALTERNATIVE I. One could break the presumed correlation between simplicity and generality of rules, thereby rejecting the claim that children posit maximally general rules. This is the solution we will focus on in this paper, but as we will show, it entails making some pretty radical proposals about the properties of grammars; this approach is NOT compatible with familiar theoretical frameworks such as Government Binding Theory or Generalized Phrase Structure Grammar.

ALTERNATIVE II. One could assume that children WOULD formulate maximally general rules except that they have data -- negative data -- that forces them to restrict their rules, even though this means complicating the mental grammar. We are not going to pursue this alternative. We don't want to claim that there are absolutely no negative data for language acquisition; nor do we want to deny the possibility of some substitute for negative data. But there are interesting reasons why these sources of information don't provide a solution for the particular kind of overgeneralization puzzle that we are concerned with here. (For discussion, see Fodor and Crain, in press.)

ALTERNATIVE III. One could assume that children do formulate maximally general rules, but that the effects of these rules are held in check by constraints on their application or their output. This is the standard way of coping with the problem of partial generalizations within a transformational theory. In the GB framework, for example, the rule responsible for topicalization (and for all other kinds of movement) is Move alpha, which is maximally general; it says that any phrase can move from anywhere to anywhere else in a phrase marker. The result would be a language with totally free word order -- except that the operation of this rule is greatly restricted by constraints like Subjacency and/or the Empty Category Principle, the binding principles, the case filter, the theta criterion and so forth.

Note that, unless this approach III is combined with approach II, in which negative data is available, all of these necessary constraints must be innate. A constraint is a negative condition on a language, and a negative condition cannot be learned from positive data. This creates a research situation that is very challenging. We assume children formulate maximally general rules. This commits us to the existence of constraints which must all be innate. If they are innate they must be universal. So we look at other languages to see if the constraints apply universally. Sometimes they do and all is well. But sometimes they don't, and then we have to make further assumptions, such as that innate constraints can be overridden by positive data; or else we have to find ways of building parameters into the innate constraints, to leave room for just the right amount of cross-linguistic variation.

We are not going to pursue this approach here, either, because it is perfectly familiar already -- you can play it for yourself any time. We are quite prepared to admit that it might turn out to be true. But there are several reasons for not embracing it without giving some thought to the alternatives. We can't go into all of our reasons here, so we will just make one quick point. The chain of argument that leads to this innate constraint approach may seem to be inexorable, but it is not. It rests crucially on the initial premise that as rules get more simple they get more general in application. And there is nothing necessary about this. It happens to be true for some kinds of linguistic rules, but it also happens not to be true for others. So it is a pretty shaky foundation to rest one's theory on.

Alternative I rejects this standard premise, and that is what makes it interesting, worth exploring. Unlike the standard approach, with its overgeneral rules held in check by innate constraints, the goal of this approach is to pitch the generalizing tendency of the rule system just right in the first place, to pitch it so that it exactly matches the extent of the generalizations that children formulate and that adult languages exhibit. If this can be done, it will obviously represent the sharpest, most explanatory account of WHY linguistic generalizations go just as far as they do and no further.

So now let's get down to work. Our goal is to see what sort of theory of language we arrive at if we DON'T use the kinds of rules for which greater simplicity means greater generality. This turns out to be a pretty drastic step, for it immediately excludes transformational rules, as well as all other

context sensitive rules. A context-sensitive rule can be simplified by omitting or reducing its context restrictions, with the result that it applies more generally. So context-sensitive rules are forbidden, so are rules which can contain features, and so are rules in which variables can appear in place of specific category designations. Each of these formal devices results in the troublesome positive correlation between simplicity and generality, and so must be rejected.

Fortunately we don't have to reject everything. Context free phrase structure rules of the old-fashioned pre-transformational variety are perfectly acceptable. Consider the rule VP --> V . NP. This encodes very specific information -- a verb is immediately dominated by a VP node and has an NP node as its right sister. We can't reduce it to anything less specific, such as: a constituent is dominated by some node and has something or other as its sister. This could only be said with variables, and standard phrase structure rules don't use variables. The simplest rule that can be formulated is thus very specific. This one handles just transitive verb-initial verb phrases; any further coverage would require greater complexity of the rule system. To handle intransitive verb phrases, for example, we would have to have another rule VP --> V, or else we would have to insert parentheses in the first rule. Verb-final verb phrases would call for yet another rule. And so on.

This fact, that phrase structure rules are inherently limited in their generalizing capacity, means that a child could learn the phrase structure rules for the language, always selecting the simplest rule compatible with the available data, and yet not wildly overgeneralize. The formal properties of the rule system force children to learn the language in small bites, one construction at a time, each one motivated by positive data. And as long as they do that, there will be no need for either negative data or innate constraints to hold them back.

There are two obvious objections to this picture, which have to be answered. First: a child's grammar will tend to undergeneralize at early stages, and it will keep expanding as the child learns more. Does this imply that the grammar will have to go on expanding forever, that the child will never really know the language? Fortunately it doesn't, for as soon as recursive rules are acquired, the grammar will suddenly take off on its own; now it will generate far more constructions than were exemplified in the data. A rule that introduces a

clause within a clause will also introduce a clause within a clause within a clause within a clause, without limit. But what's interesting about THIS kind of leaping ahead of the data is that it seems to be exactly what children do do, and hence exactly what we would want our theory to predict. We are trying to tailor the generalizing tendency of rules to the partial generalizing tendency of children, and here we seem to be succeeding.

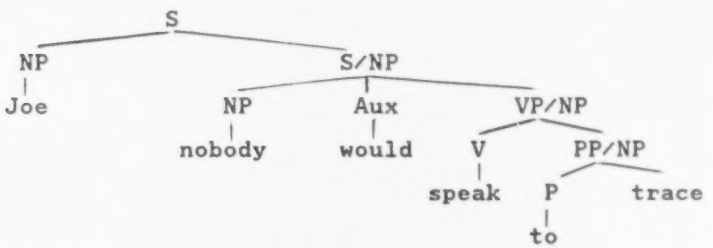
So: we are proposing to avoid overgeneralization by limiting the grammar to context free phrase structure rules. But now we must face the second possible objection, which is that we have swung the pendulum too far in the opposite direction, that we are stuck with a system which UNDERgeneralizes so badly that it can't capture the facts of English or any other natural language. Didn't Chomsky prove back in 1957 that context free phrase structure grammars are too weak to account for natural languages? The answer is that he did, but only for a very restricted kind of phrase structure grammar. All sorts of enrichments of transitional phrase structure grammar have been proposed in the last few years, particularly within the theory called Generalized Phrase Structure Grammar (GPSG), and there is no reason why we shouldn't take advantage of these new formal devices -- as long as we take care that, in providing the extra descriptive power that is needed, these enrichments don't swing us right back again into overgeneralization.

We will argue that there IS a satisfactory middle path, and that it results from accepting the 'slash categories' of GPSG while rejecting the metarules that generalize across them. (In more recent versions of the theory, the work of the metarules is done by general feature instantiation principles instead. But the consequences for acquisition are equally disastrous, so we would have to reject these too. To keep the exposition simple, we will consider here just the older, metarule version of the theory.)

What follows is a brief tutorial on GPSG for those who have never encountered it before. What a traditional phrase structure grammar couldn't do was relate two positions in a phrase marker that are distant from each other (and are not sisters). But it is just such relations that we find in natural language sentences, e.g., a relation between a moved phrase (such as a topicalized noun phrase) at the front of a sentence, and its trace (or the deep structure position it was moved from) down at the other end of the sentence. That's why Chomsky added transformational rules to phrase structure rules: transformations

CAN link two distant positions. However, phrase structure rules can too, if we set them up so that they establish a path through the tree structure, linking one position to the other. Consider the phrase marker (4).

(4)



It has a chain of special nodes, marked with a slash followed by an NP, running down through the tree. The initial noun phrase Joe is topicalized, and the path of slashed nodes runs from the S node next to it, to the trace that marks its 'deep' position in the sentence. By annotating the nodes of the tree with these slash symbols, we connect the two positions and can establish a dependency between them. So now we can insure (just as a transformation can) that we won't get a topicalized noun phrase at the front without a trace later on, and we won't get a trace later on without a topicalized noun phrase at the front.

The rules needed to generate (4) are shown in (5).

(5) S --> NP . S/NP
S/NP --> NP . Aux . VP/NP
VP/NP --> V . PP/NP
PP/NP --> P . trace

Observe that even though they contain slash categories, these are perfectly ordinary phrase structure rules. Thus, phrase structure rules enriched with slashes are powerful enough to simulate the effect of a transformation, and yet they are still inherently undergeneralizing and thereby solve the language acquisition problem. However -- and this is a big however -- the grammar for a natural language such as English would have to contain an enormous number of such rules. It will contain the usual basic phrase structure rules, the additional rules with slashes listed above, and a great many other such rules for other kinds of extraction in English. The rules in (5) only permit topicalization of a noun phrase from inside a prepositional phrase in the main clause. To topicalize a direct object, another slash rule would be needed; to topicalize out of subordinate clauses we would need more, to topicalize prepositional phrases we would need yet more. Unless we can somehow streamline its format, the grammar would be immense; and worse still, it would miss some

obvious generalizations, such as that each rule with slashes in it is identical to an ordinary basic phrase structure rule except for the presence of the slash annotations on its nodes.

This is what GPSG uses metarules (or feature instantiation principles) for -- to capture the generalizations that are beyond the capacity of a pure phrase structure grammar (with or without slashes). Sentences are generated by a context free phrase structure grammar, just as we have described, but this phrase structure grammar is generated by a metagrammar. Then the fact that there are a great many phrase structure rules is no longer a problem; the rules are not being listed any more, they are being generated. And the missing generalizations can now be captured, not in the phrase structure grammar, of course, but at the level of the metarules which generate the phrase structure grammar. For example, the metarules can capture the correspondence between slash category rules and basic rules if they generate the slash category rules FROM the basic rules -- they take a basic rule and add slash annotations to it. That is what the metarule in (6) is doing.

(6) X --> Y . Z ==> X/W --> Y/W . Z
X/W --> Y . Z/W

The input of the metarule is a basic rule; its output is the corresponding slash category rules.

This is all very clever, but unfortunately it does lead straight back into overgeneralization. We are to suppose that a child either learns or knows innately the metarule (6). As the child learns new basic phrase structure rules, the metarule will apply to them and thereby generate slash category rules. The danger, of course, is that the metarule will generate too many slash category rules. In the worst case, if the metarule produced all possible slash rules, it would permit extraction of any phrase from anywhere, just as Move alpha does, and with the same appalling consequences for acquisition. On the other hand, if the metarule doesn't generate too many slash category rules, that must be because it is subject to some constraints, and these constraints must be either innate and universal, or else must be learned in violation of the least effort principle and the lack of negative data. To cut a long story short, the metarule for slash categories runs into exactly the same problems as the transformation Move alpha does. The phrase structure rules in themselves express very modest generalizations, but the metarule exhibits the same kind of

runaway overgeneralization as a transformation does; so, just like a transformation, it would have to be patched up with innate constraints in order to account for the observed preference for partial generalizations.

To satisfy ALTERNATIVE I, therefore, GPSG metarules must be jettisoned along with transformations. But we have at least made some progress by borrowing from GPSG, for the slash category rules that we have adopted allow us to handle long distance dependencies that were beyond the power of more traditional phrase structure grammars. All we are short of now is some SAFE way of cutting the huge set of phrase structure rules down to size. But in fact a mechanism for doing this has been around for years -- the so-called collapsing conventions or abbreviatory devices.

In (7), we have collapsed three rules into a single rule schema, using angle brackets. (We are assuming here a convention that the contents of a pair of angle brackets may be omitted just in case they are identical to the contents of the associated pair of angles.)

(7) S $\langle$/ $\{$NP / PP$\}$ $\rangle$ --> NP . VP $\langle$ $\rangle$ | equivalent to: S --> NP . VP
S/NP --> NP . VP/NP
S/PP --> NP . VP/PP

The collapsed rule says that S can be expanded as NP . VP, with either no slashes, or NP slashes or PP slashes in the appropriate places. The status of these angle brackets with respect to the simplicity metric and acquisition is exactly like that of the more familiar parentheses in (8).

(8) VP --> V . (NP) | equivalent to: VP --> V . NP
VP --> V

For clarity, we will illustrate the general point with reference to parentheses first.

In terms of formal simplicity, the rule schema with parentheses is simpler than two separate independent rules. But the rule schema with parentheses is more complex than either rule alone. Now consider what this means for acquisition. Imagine a child who has encountered only transitive verb phrases so far. (Note this is not meant to be a realistic example.) The child will formulate the rule VP --> V . NP, but will not put parentheses around the NP, for that would be a more complex rule than is needed to cover current data. Therefore the child won't predict the existence of intransitive verb phrases in advance of relevant data, and so won't be in any danger of OVERpredicting,

overgeneralizing. On the other hand, when an intransitive verb phrase is encountered, it can be added to the grammar very cheaply, at the cost of just a pair of parentheses in a rule the child already has. Thus learners get all the benefits of generalization (simplicity of the grammar, significant relationships captured) but it is post hoc generalization, not predictive generalization. It follows safely along behind children's data; it doesn't push them forward recklessly in advance of it.

In exactly the same way, the use of angle brackets to collapse a variety of slash rules permits post hoc generalizations about extraction. A child can learn piecemeal, one kind of extraction at a time, never going beyond what is evidenced in the available data; but as all the other extraction possibilities are eventually encountered, the learner will be able to encode them economically in the mental grammar and capture the similarities among them.

Our conclusion is that it IS possible to devise a theory of grammars which favors generalization without overgeneralization in language acquisition. We have not tried to argue here that this nonstandard theory is correct (see Fodor and Crain, in preparation, for some problems and puzzles), but we do commend it as a valuable basis for research. It should serve as the benchmark against which more elaborate theories of innate linguistic knowledge are judged.

REFERENCES

Bowerman, M. (1974) Learning the structure of causative verbs: A study in the relationship of cognitive, semantic and syntactic development. In E. Clark (ed.), Papers and Reports on Child Language Development, No. 8, 142-178. Stanford University Committee on Linguistics.

Fodor, J.D. and S. Crain (in press) Simplicity and generality of rules in language acquisition. In B. MacWhinney (ed.), Mechanisms of Language Acquisition. Proceedings of the 20th Annual Carnegie-Mellon Conference on Cognition. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Fodor, J.D. and S. Crain (in preparation) On the Form of Innate Linguistic Knowledge. To be published by Bradford Books, Cambridge, Mass.

Halle, M. (1962) Phonology in generative grammar. Word 18, 54-72.